FAQs:

How does this tool account for small numbers?

- The tool uses a Bayesian hierarchical model that borrows strength from priors, neighboring areas and over time to produce estimates for small areas. The estimates that are produced from this model are estimates that represent an *approximated* value based on modeled data.
    - In states with low populations of non-white individuals, estimates for race stratified age-adjusted rates are a little more complex. For example, most of Iowa has a very small population of non-white individuals, so most of the state has 0 cancer cases when stratified by race. For these areas, the raw age-adjusted rate is 0. The Bayesian model underlying the estimates of the tool works by borrowing strength from neighboring areas. When stratified by race, the neighboring areas also have 0 cases. The model then defaults to the state race-specific average. In summary, all of the areas that show state average are probably actually 0, meaning the state average is not the actual age-adjusted rate in those regions, but there is not enough evidence to suggest it is any different from the state average because there is no data.

What calculations are done to get these results?

Please see below for detailed calculations of the estimates produced by this tool.

- *Age-adjusted rate:* Cases per 100,000 people were estimated for each age group within a location (i.e., ZCTA or county) and time period. The age groups are under 40, 40-49, 50-59, 60-69, 70-79, and 80 and up. The rate for each age group was multiplied by standard population weights based on the 2010 U.S. Census. The weighted rates were then added together to create a single age-adjusted rate. Our team used the 2010 Census data and used US weights versus state-specific weights.
- *Risk probability (overall):* One state-adjusted rate was calculated for the 2004-2018 time period. Similarly to the risk probability within group calculation, each age-adjusted rate estimate by ZCTA and county level was compared to the state age-adjusted rate. Then as before, these were averaged by county and ZCTA level to come up with a probability that the county or ZCTA age adjusted rate is greater than the state-adjusted rate for the 2004-2018 time period.
- *Prevalence vs incidence:* Incidence measures the number of new cancer cases in a certain period of time, while prevalence includes all existing cancer cases in a certain period of time.

Where does the data come from?

- The cancer data comes from Surveillance, Epidemiology, and End Results (SEER) Program collected data. Estimates for mortality, incidence, and late-stage incidence are produced from this data using our Bayesian hierarchical modeling.
- Census data is used for the population data to calculate the age-adjusted rates for incidence, mortality, and late-stage incidence. Census data is also used for population density estimates.

How are the data collected?

- For details on SEER data collection, please see the SEER website: https://seer.cancer.gov/about/overview.html

Can I export data/maps from this tool?

- Both estimate tables and maps are exportable from CAMSA. Please see the How To Guides for exporting tables and maps.

What can I do about high cancer incidence in my area?

- Reducing cancer rates is a complex problem that will require complex and collaborative solutions. The Iowa Cancer Plan is a helpful tool that provides a guide for ways all Iowans can work together to reduce our collective cancer risk. If you have ideas or questions about how to address high cancer incidence in your community and are interested in support from the University of Iowa, please reach out ICR-Research@uiowa.edu

How often is CAMSA updated?

- CAMSA will be updated annually as data from cancer registries are released.

I have a question, comment, or feedback on the CAMSA tool. Who can I contact?

- Please email us at ICR-Research@uiowa.edu

How has this tool been used?

- Please see below for publications using this tool:
    -

How is this tool different than other tools?

- Existing tools, like Cancer InFocus, NCI Cancer Zones, and CDC Places, have similar use cases, but different features make CAMSA uniquely useful for surveying small area cancer data. Cancer InFocus and NCI Cancer Zones provide data at the county level or larger, and while CDC Places provides data at the ZCTA and census tract levels, it does not have cancer-type data available. Our project combines the best features from each of these existing tools, while employing novel modeling methods to produce estimates for eight leading cancer types.

How do I interpret an "estimate"?

- There are different ways to interpret different estimates. Please use the interpretation suggested on the map as a guideline for interpretation.

What does uncertainty mean? How is it calculated?

- The 'uncertainty' layer represents a classification of the standard deviation of the age-adjusted rate (AAR) into three levels—low, medium, and high. This shows how confident we are in estimating the age-adjusted rate for each unit. Low uncertainty means the estimate is more reliable, while high uncertainty means there is more variation in the data. These levels are determined by ranking all standard deviations of the age-adjusted rate from highest to lowest and dividing them into three equal groups. The first group, with the highest standard deviations, represents high uncertainty. The second group, with moderate values, represents medium uncertainty. The third group, with the lowest standard deviations, represents low uncertainty.

Why would certain rates be smaller than others?

- Cancer is a very complex set of diseases that have many causes, including environmental, genetic, and lifestyle factors. It is difficult to identify the causes of cancer for a certain person, but there are certain exposures that we know increase risk at the community level. For example, eating a healthful diet, engaging in physical activity, and not using tobacco (smoking) can substantially reduce risk, as can getting age-appropriate screenings. If your community has higher (or lower) cancer rates, there could be many contributors. If you'd like to talk to someone about factors that may be related to cancer rates in your community, please email us at ICR-research@uiowa.edu

How would estimates change by doing census tracts rather than ZCTAs?

- The same statistical model would be used with a different structure of the neighborhoods. Census tracts are being considered for future iterations of the tool.

What is statistical influence of neighboring states on border counties/ZCTAs?

- Neighboring states do not influence border counties or ZCTAs because we do not have data from neighboring states. Our analyses assume each state is an island; we can add in uncertainty to account for those edges, but it would not impact estimates greatly.

Glossary of terms

- Risk Probability: Measure of elevated risk for a location after adjusting for age. Values closer to 1 signify higher risk while values closer to 0 mean lower risk for the location compared to overall state risk
- Age-adjusted rate: Estimated cases per 100,000 people, adjusted for the ages of individuals in the location
- Population density: The population count from the 2010 U.S. Census, stratified by sex or race/ethnicity, divided by the area measured in units of 1,000 km^2.
- ZCTA: ZIP Code Tabulation Areas – ZCTAs are based on census blocks so they are slightly different than ZIP codes, though there is a great amount of overlap between ZIP Codes and ZCTAs. ZCTAs are representative of geographic locations of populated areas, so ZCTAs will not necessarily exist for ZIP Code areas with only businesses, for single or very few addresses, or for large unpopulated areas. Because ZCTAs are based on the most recent Census, they are more stable than ZIP Codes and do not change as frequently
- Incidence rate: Incidence rate is a measure of the number of new cases of a specific cancer in a geographical area over a specific period of time.
- Late-stage incidence: The number of new cancer cases diagnosed at a more advanced stage in a specific geographical area over a specified period of time. What is defined as late-stage differs for each cancer type. Data included here are based on SEER definitions of late-stage for each cancer type.
- Uncertainty: This shows how confident we are in estimating the age-adjusted rate for each unit. It is divided into three levels: low, medium, and high uncertainty. Low uncertainty means the estimate is more reliable, while high uncertainty means there is more variation in the data. These levels are determined by ranking all standard deviations of the age-adjusted rate from

highest to lowest and dividing them into three equal groups. The first group, with the highest standard deviations, represents high uncertainty. The second group, with moderate values, represents medium uncertainty. The third group, with the lowest standard deviations, represents low uncertainty.